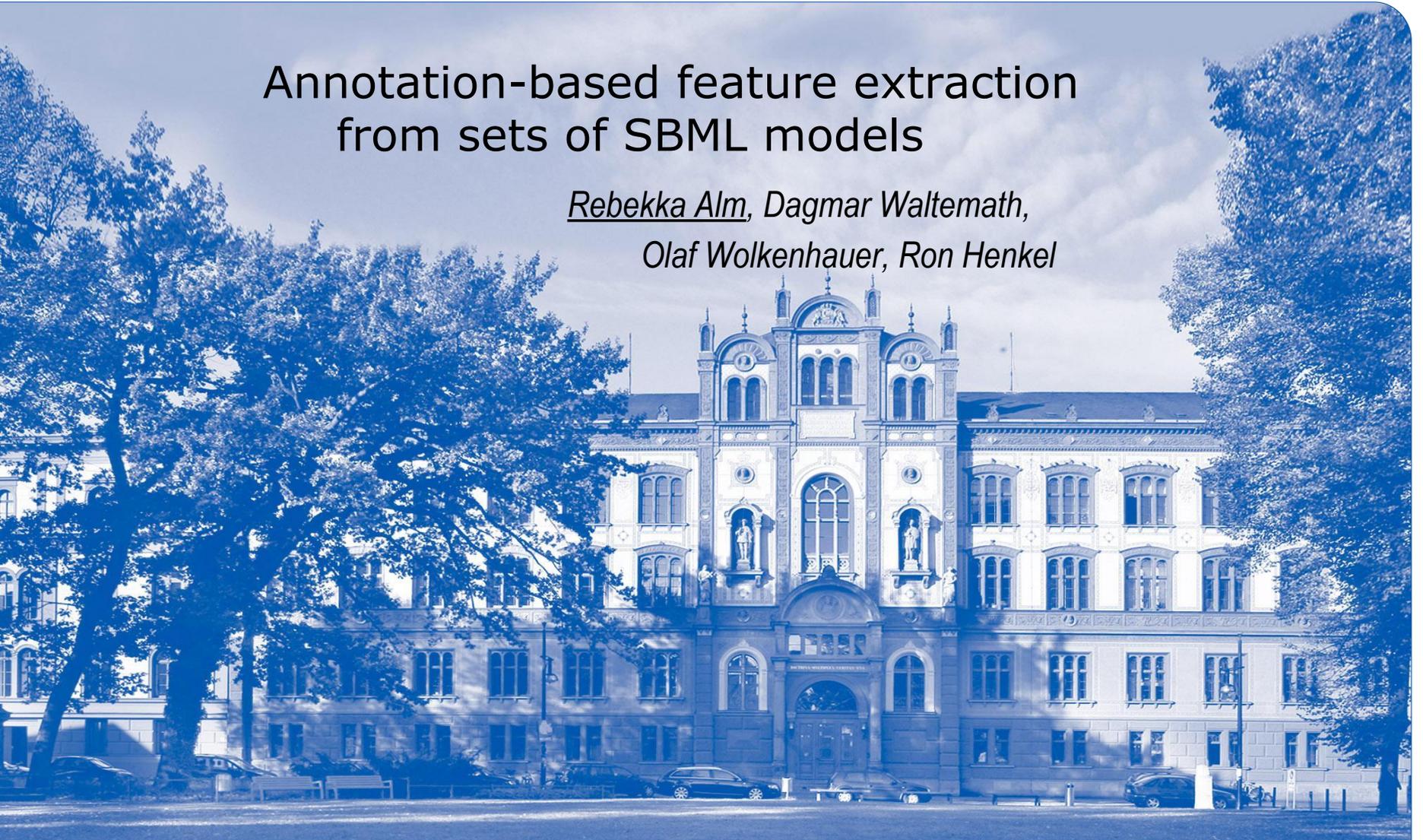
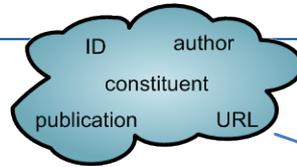


Annotation-based feature extraction from sets of SBML models

*Rebekka Alm, Dagmar Waltemath,
Olaf Wolkenhauer, Ron Henkel*



Motivation: BioModels retrieval



annotation index

BM-ID	species	speciesURI	...
BIOMD00000241	C_p	Chebi:27732 Kegg:C07481	...



ranked
model
retrieval



bio-ontology

ranked results



other resources

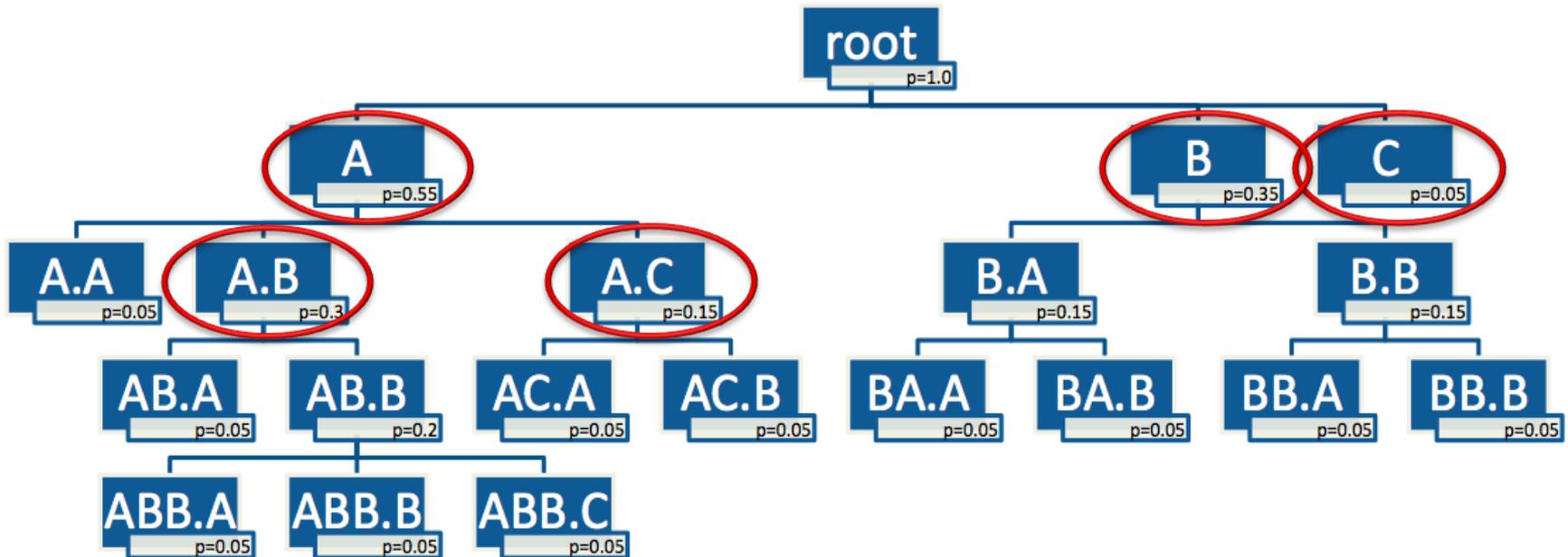
Rank	Score	BioModels ID	Name
1.	9.3822	BIOMD0000000241	Shi2003_Caffeine_pressor_tolerance
2.	0.5000	BIOMD0000000015	Curto1998_purineMetabol
3.	0.1495	BIOMD0000000017	Hoefnagel2002_PyruvateBranches

#	dimension	importance	description
1	Administrative	none	ids, file name, version, formalism
2	Person	medium	creator, encoder, submitter, publication author
3	Dates	low	creation and modification date
4	Publication	high	title, abstract, full-text, journal
5	Constituents	very high	compartment, species, reaction
6	User generated	very high	keywords, tags, remarks, changes

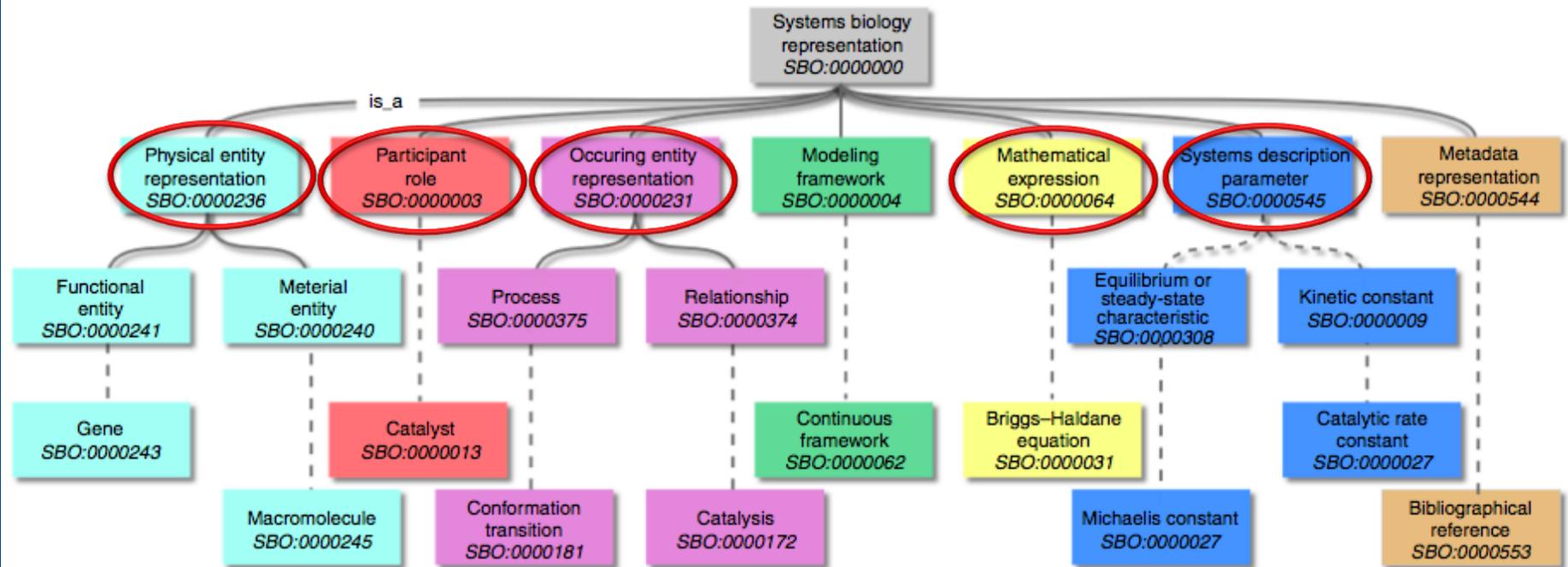


How can we extract model features that characterize model sets?

Choosing representatives

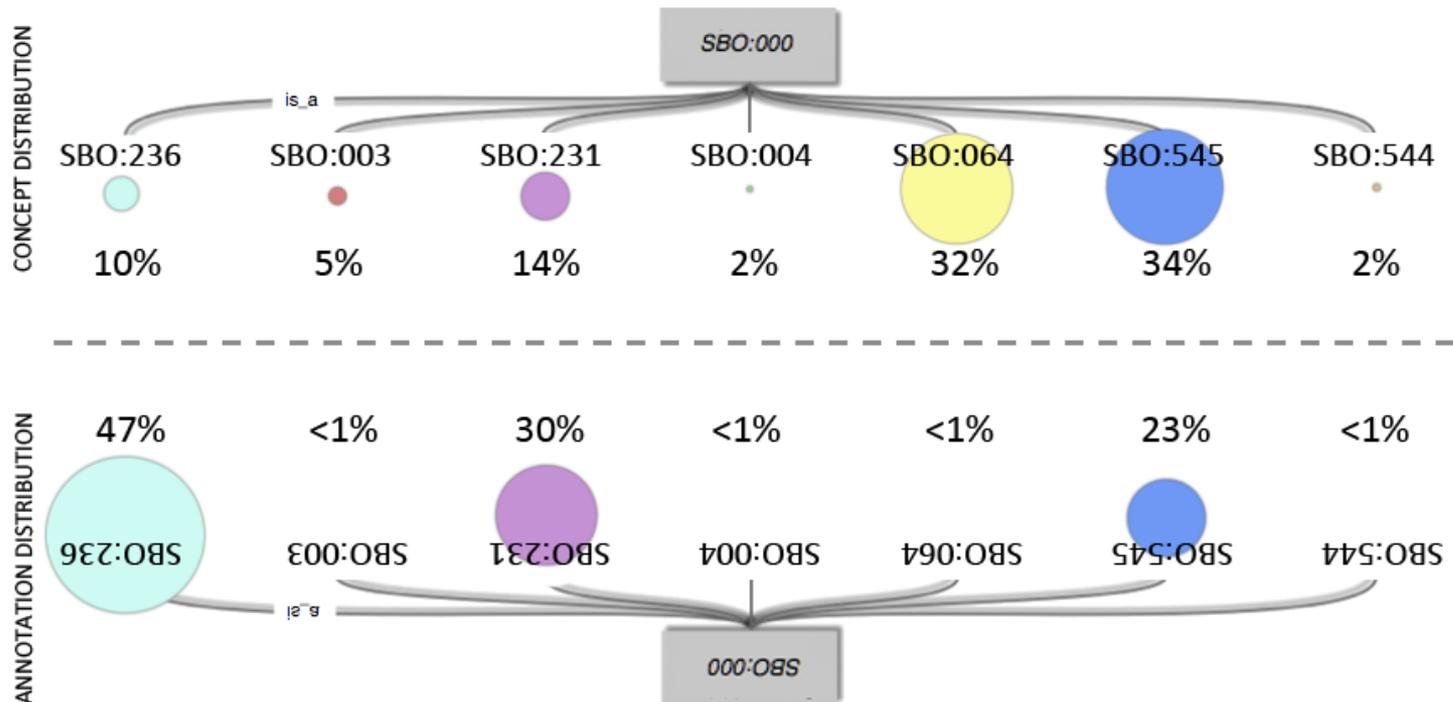


Systems Biology Ontology (SBO)



Source: Courtot et al.: Controlled vocabularies and semantics in systems biology. In: Molecular Systems Biology 7 (2011), Nr. 1.

Problem: Terms are unequally distributed



SBO Terms are shortend (e.g. SBO:004 equals SBO:0000004):

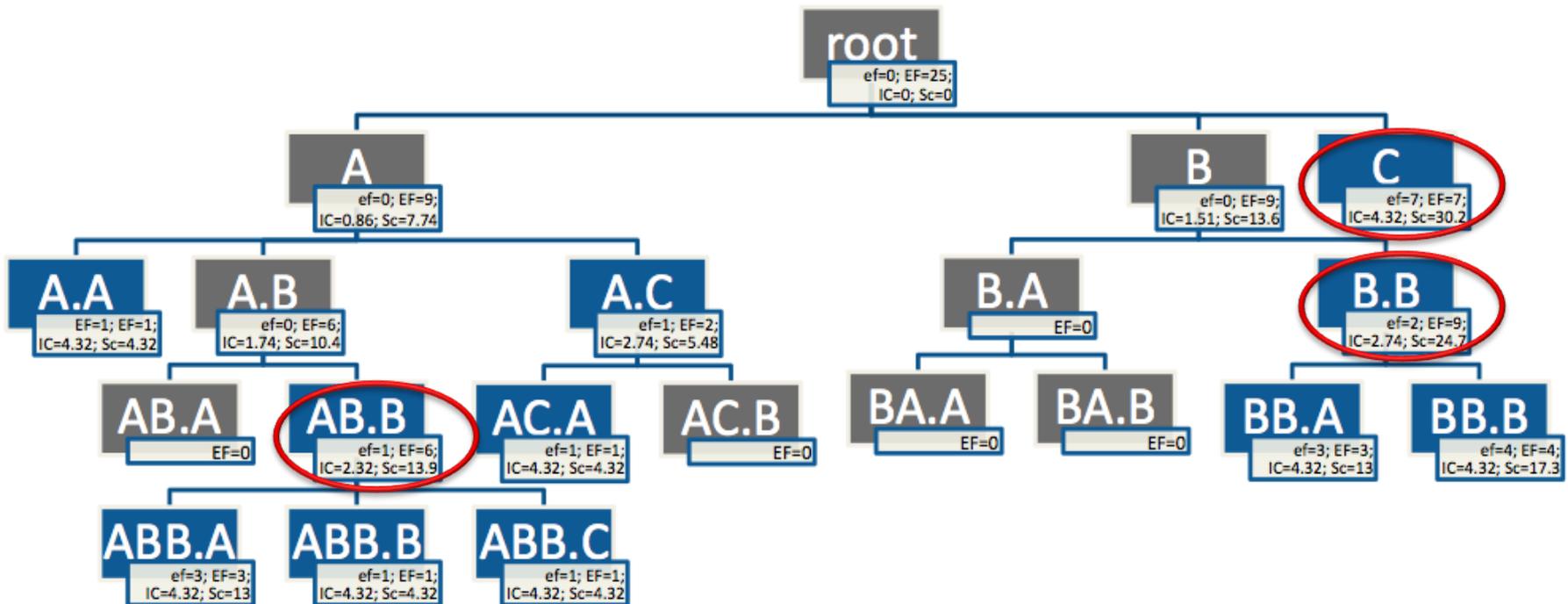
SBO:000 Systems biology representation; **SBO:236** Physical entity representation; **SBO:003** Participant role;

SBO:231 Occuring entity representation; **SBO:004** Modeling framework; **SBO:064** Mathematical expression;

SBO:545 System description parameter; **SBO:544** Metadata representation

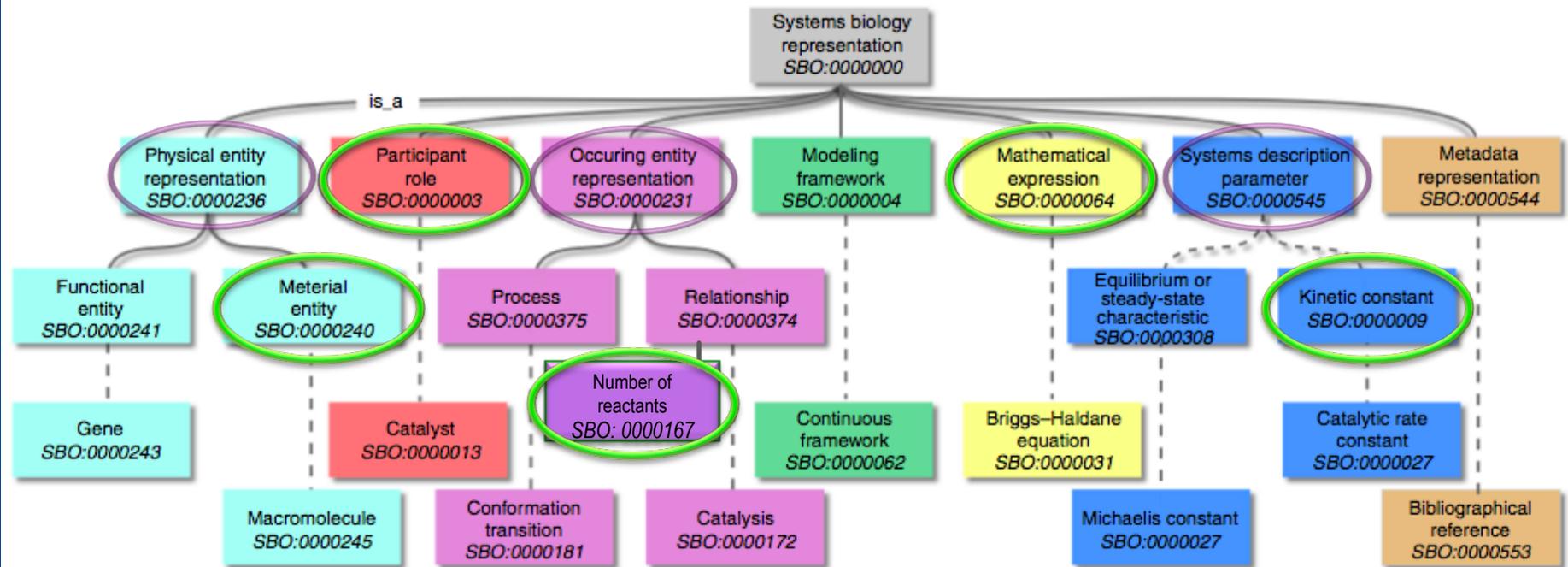
BioModels Database Release 25 (490 files)

Choosing representatives



Information content $IC(c) = -\log p(c)$
 Score $Sc(c) = IC(c) * EF(c)$

SBO: Updated features



Source: Courtot et al.: Controlled vocabularies and semantics in systems biology. In: Molecular Systems Biology 7 (2011), Nr. 1.

Possible Solutions

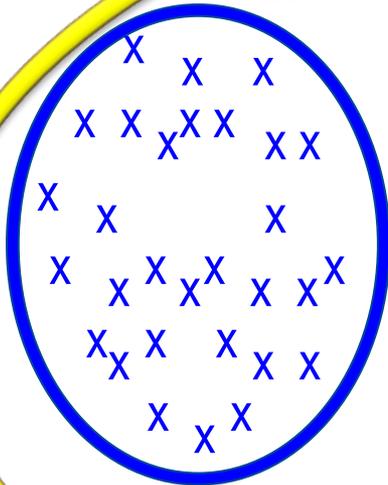
Method	Direction	Knowledge Base	Score Value
1	Top-down	Ontology	p (<i>concept probability</i>)
2	Top-down	Ontology and annotated models	EF (<i>aggregated entity frequency</i>)
3	Bottom-up	Ontology and annotated models	EF (<i>aggregated entity frequency</i>)
4	Bottom-up	Ontology and annotated models	EF, IC (<i>aggregated entity frequency, information content</i>)

Test Data

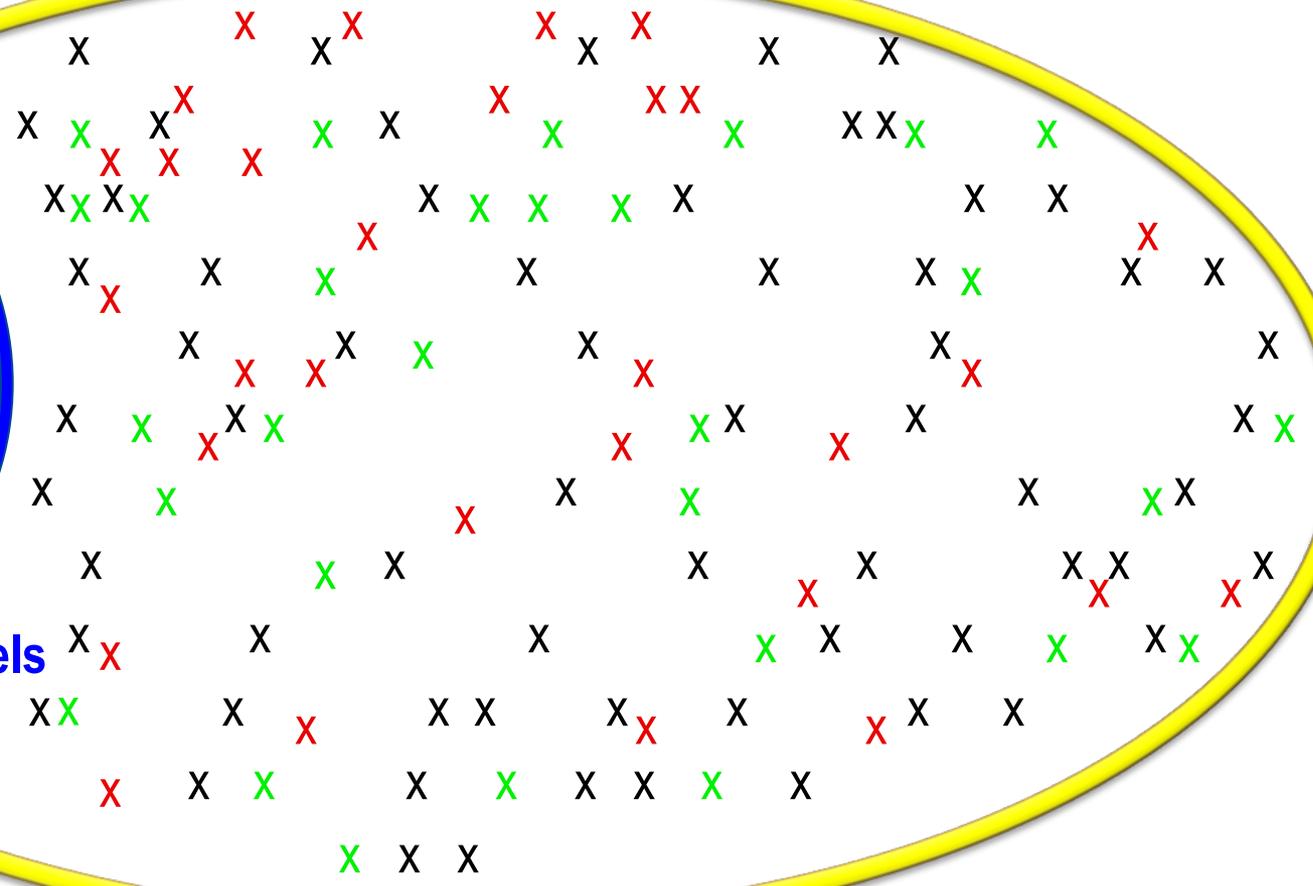
- SBML models from BioModels Database
- Test sets:
 - BioModels Database
 - Random sets
 - Cell Cycle models
- References to bio-ontologies (all models)
 - ~11,000 references to GO concepts
 - ~ 6,000 references to ChEBI concepts
 - ~13,000 references to SBO concepts

* BioModels Database Release 25 (490 files): <http://www.ebi.ac.uk/biomodels/>

Models in BioModels Database

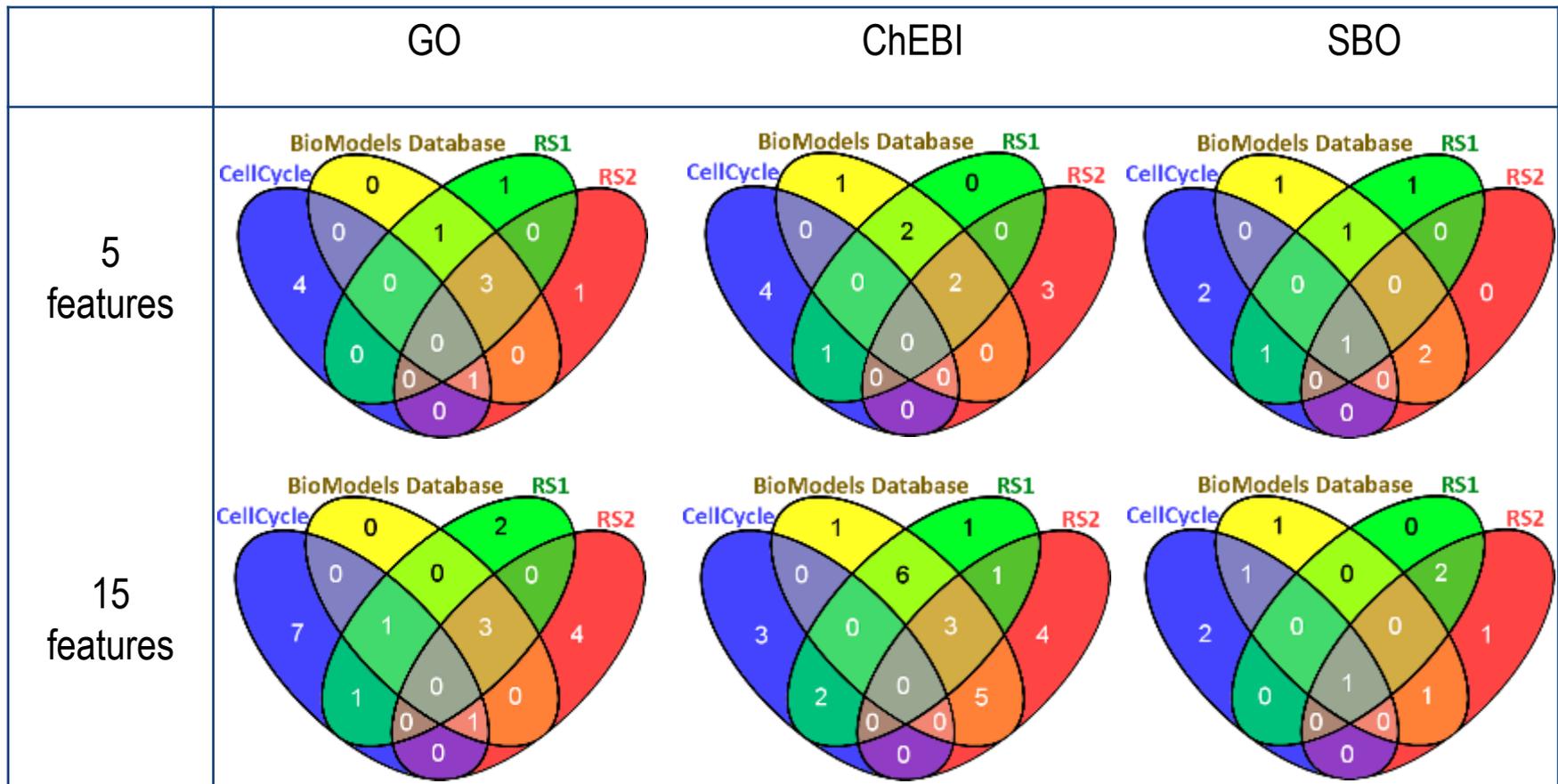


CellCycle Models



BioModels Database Release 25 (490 files)

Features identified by Method 4



Results

Features for CellCycle model set, GO, Method 4, 5 features:

- GO:0022411 (depth 4) cellular component disassembly
- GO:0030163 (depth 5) protein catabolic process
- GO:0051726 (depth 5) **regulation of cell cycle**
- GO:0065009 (depth 3) regulation of molecular function
- GO:0071822 (depth 5) protein complex subunit organization

- Average depth:
 - BioModels Database: 2.6
 - CellCycle models: 4.4

Evaluation: Similarity between model sets

- Calculated based on the similarity of their characterized features
- Ontology: Gene Ontology
- Approach: Method 4

Model sets	5 Features	15 Features
BMDB vs. RS1	0.84	0.59
BMDB vs. CellCycle	0.40	0.30
CellCycle vs. RS1	0.25	0.36



Summary

- Four feature extraction approaches
- Specific and distinct features, not overgeneralized
- Similarity value of sets of models
- Format agnostic and expandable
- Comparison between sets of models of arbitrary formats

Thank you for your attention.



SEMS
simulation experiment management system
sems.uni-rostock.de

HARMONY 2015
= Wittenberg =

VW Summer school 2015 =
Rostock =

Rebekka Alm
Visual Computing Research and
Innovation Center
Fraunhofer IGD Rostock
rebekka.alm@igd-r.fraunhofer.de
www.igd-r.fraunhofer.de

Implementation

